Guest Editorial for the Special Issue on Fuzzy Rough Sets for Big Data

Recent advances in computing technology imply collecting vast amount of data coming from various sources, such as the Internet, senor monitoring systems, social networks, mobile communication systems, transportation systems, and so on. We continue to encounter an explosive growth in Big Data coming with highly visible aspects of Volume, Variety, Velocity, Veracity and Value. These 'Five Vs' are the key features of big data; they also come with inherent uncertainties when it comes to their representation, processing, and analysis[1]. Big Data have emerged as a hot topic along their promising benefits in many fields. As such data often contain a significant amount of unstructured, uncertain and imprecise data, which are structurally complex and incomplete, traditional techniques and methods of data analysis are less effective when being considered in large-scale incomplete information system to extract knowledge. Installation of various sensors makes it possible for numerous variables to be captured and in turn, this leads to an exponential growth of the problem size to be handled. Real-life big data often refer to incompleteness, uncertainty and vagueness. The conventional techniques of knowledge discovery, ranging from models, algorithms, and systems, have been challenged on how to store, manage, process, and analyze big data.

In the recent past, the evolution of research interest has focused on fuzzy sets and rough sets. They provide two powerful conceptual and algorithmic vehicles for multiple-view data analysis, which is of vital importance for understanding of data being completed at different granularity levels[2]. Fuzzy rough sets have emerged as a new synergistic platform delivering the advantages of two complementary areas (fuzzy sets and rough sets). In the rough case, a set is provided with a lower approximation and an upper approximation, which allow for a granular representation of knowledge and an excellent description of the uncertain region. In the fuzzy case, relations between objects and sets or relations among objects are characterized by degrees of membership. This allows for great flexibility in dealing with imprecise information. Fuzzy rough sets is considered to be a more powerful model for analyzing uncertainty in big data, and it can be directly applied to a variety of knowledge analytical problems of large-scale datasets. Compared with the classic rough set theory, fuzzy logic operators are used to replace the equivalence relation. In the fuzzy rough case, elements have a membership in the range, which allows for a greater flexibility in handling uncertain information. It has been a timely trend to address the classical and new-emerging data mining problems in big data. Fuzzy rough set theory is powerful in handling uncertainty and has been extensively applied in the fields of feature selection, reasoning with uncertainty, and classification learning[3][4]. So there is a sound potential to improve reasoning and understanding of big data by using the fuzzy rough sets methods.

The sixteen articles forming this special issue of the IEEE TRANSACTIONS ON FUZZY SYSTEMS represent some of the latest developments in the state-of-the-art of fuzzy rough sets in big data and knowledge processing. These papers were rigorously evaluated according to the normal strict reviewing process. The objective of this special issue devoted to provide a broad forum, where professionals coming from various scientific, engineering and application communities can exchange their recent ideas and accomplishments and present ensuing models and algorithms along with a thorough discussion on their performance in practical applications. The contents of these sixteen articles are briefly described below.

The paper entitled "COEVIL: A Co-Evolutionary Model for Crime Inference Based on Fuzzy Rough Feature Selection," by Liu *et al.*, comprehensively investigates the reliability and applicability of proposing a co-evolutionary model to formulate the interaction pattern among the crimes and locations, and develops a fuzzy rough set-based method to find the good feature subset to deal with the challenges of uncertain and vagueness by the large-scale crime data.

The paper entitled "Fuzzy Rough Attribute Reduction for Categorical Data," by Wang *et al.*, proposes a new fuzzy rough set model for big categorical data by introducing a variable parameter to control the similarity of samples. This model employs the iterative computation strategy to define fuzzy rough approximations and dependency function and it can better characterize the similarity of categorical big data than crisp equivalence relations.

The paper entitled "Hypotheses Analysis and Assessment in Counter-terrorism Activities: A Method based on OWA and Fuzzy Probabilistic Rough Sets," by Fujita *et al.*, combines probability, fuzzy and rough set theories and presents a new interactive method to assess hypotheses for counter-terrorism activities in the analysis of intelligence information by using behavioural models of known terrorist groups.

The paper entitled "Distributed Feature Selection for Big Data using Fuzzy Rough Sets," by Kong *et al.*, proposes the novel Distributed Fuzzy Rough Set (DFRS) based feature selection. They specially design a dynamic data decomposition algorithm to divide the dataset into subdatasets and assign them to distributed nodes in cloud, with maintaining the global information on each distributed node without conserving the entire fuzzy relation matrix.

The paper entitled "Novel Incremental Algorithms for Attribute Reduction from Dynamic Decision Tables using Hybrid Filter–Wrapper with Fuzzy Partition Distance," by Giang *et al.*, presents new incremental methods using the hybrid filter wrapper with fuzzy partition distance on fuzzy rough set to deal with the dynamic decision tables of big data.

The filter stage determines the candidates for reduction, and the wrapper stage finds the reduction with the best classification accuracy.

The paper entitled "Fast and Scalable Approaches to Accelerate the Fuzzy k Nearest Neighbors Classifier for Big Data," by Maillo *et al*., develops a fuzzy k-nearest neighbors (FkNN) model capable of handling large datasets accurately and quickly. They present the global approximate hybrid spill tree FkNN and local hybrid spill tree FkNN, where the hybrid spill-tree model is used as the base algorithm due to its balance between scalability and accuracy.

The paper entitled "Fusing Fuzzy Monotonic Decision Trees," by Wang *et al.*, focuses on developing a fusion method based on attribute reduction to solve ordinal classification problem in big data applications. They define a discernibility matrix with fuzzy dominance rough set and introduce the algorithms for finding reductions based on it, where each reduction forms a feature subspace with original information.

The paper entitled "Active Incremental Feature Selection Using A Fuzzy Rough Set-Based Information Entropy," by Zhang *et al*., investigates the incremental feature selection using a fuzzy rough set-based information entropy with incoming instances, and proposes an active incremental filter-wrapper approximation reduction method (AIFWAR).

The paper entitled "A Hopping Umbrella for Fuzzy Joining Data Streams from IoT Devices in the Cloud and on the Edge," by Mrozek *et al.* presents the idea of a hopping umbrella which fuzzifies timestamps from sensor readings while joining data streams from asynchronous IoT devices in a flexible way. The hopping umbrella implements the fuzzy join operation in time windows for big data.

The paper entitled "Scalable approximate FRNN-OWA classification," by Lenz *et al.*, proposes the approximative fuzzy rough nearest neighbours classification with ordered weighted averaging operators(FRNN-OWA) to calculates upper and lower approximations of decision classes by using the approximate nearest neighbours.

The paper entitled "Multiobjective Evolution of Fuzzy Rough Neural Network via Distributed Parallelism for Stock Prediction," by Cao *et al*., constructs a fuzzy rough neural network model and proposes a distributed parallel multiobjective evolutionary algorithm by introducing a more powerful optimizer as well as removing the crossover operator for the big data processing.

The paper entitled "Scalable Fuzzy Rough Set Reduct Computation Using Fuzzy Min-Max Neural Network Preprocessing," by Kumar*et al*., proposes a novel Fuzzy Min-Max Neural Network (FMNN) model to enhance the scalability of fuzzy rough sets approaches, where FMNN model is used to reconstruct the object-based decision system into fuzzy hyperboxes based interval-valued decision system.

The paper entitled "Medical Image Segmentation by Partitioning Spatially Constrained Fuzzy Approximation Spaces," by Roy *et al*., introduces a novel rough-fuzzy clustering algorithm for image segmentation, named sRFCM, which judiciously integrates the merits of rough-fuzzy clustering and local spatial constraint of each pixel.

The paper entitled "Shadowed Neighborhoods based on fuzzy rough transformation for Three-Way Classification," Yue *et al*., proposes a novel shadowed set based on fuzzy rough transformation to construct shadowed neighborhoods for uncertain big data classification and implements a three-way classification algorithm to distinguish data instances into certain classes and uncertain case.

The paper entitled "A Hierarchical Clustering Approach to Fuzzy Semantic Representation of Rare Words in Neural Machine Translation," by Yang *et al*., proposes a fuzzy semantic representation method for rare words through a hierarchical clustering method to group rare words together, and integrates it into the encoder-decoder framework on large-scale Chinese-to-English translation tasks.

The paper entitled "Optimize TSK Fuzzy Systems for Regression Problems: Mini-Batch Gradient Descent with Regularization, DropRule and AdaBound (MBGD-RDA)," by Wu *et al*., extends three powerful optimization techniques for neural networks to Takagi-Sugeno-Kang (TSK) fuzzy systems, and proposes three novel techniques specifically for training TSK fuzzy systems in uncertain big data environments.

We would like to take this opportunity to express sincere thanks to all authors who submitted papers for this issue. Our thanks also extend to the dedicated reviewers, whose expertise and critical, yet constructive comments have been indispensable in improving the quality of this Special Issue. The enthusiastic supports of the Special Issue Editor-in-Chief, Prof. Robert John, the Journal Editor-in-Chief, Prof. Jon Garibaldi and Editorial Administrator, Clair Morton are highly acknowledged. They worked closely with the Guest Editors to ensure this issue's excellence and guarantee its success. This work has been carried out in part by the National Natural Science Foundation of China under Grant Grant 61976120, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20191445, in part by the Six Talent Peaks Project of Jiangsu Province under Grant XYDXXJS-048, and sponsored by Qing Lan Project of Jiangsu Province.The support from the Canada Research Chair (CRC) and the Natural Sciences and Engineering Research Council of Canada (NSERC).

WEIPING DING*
School of Information Science and Technology
Nantong University
Nantong, 226019, China

WITOLD PEDRYCZ
Department of Electrical and Computer Engineering
University of Alberta
Edmonton, AB T6G 2R3, Canada

CHIN-TENG LIN
Faculty of Engineering and Information Technology
University of Technology Sydney
Ultimo, NSW 2007, Australia

* Corresponding author.
E-mail address: dwp9988@l63.com (W. Ding)

REFERENCES

[1] C. L. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Inf. Sci.*, vol. 275, pp. 314-347, Aug. 2014.

[2] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. Gen. Syst.*, vol. 17, pp. 191-209, 1990.

[3] A. M. Radzikowska and E. E. Kerre, "A comparative study of fuzzy rough sets," *Fuzzy Sets Syst.*, vol. 126, no.2, pp. 137-156, Mar. 2002.

[4] W. Ding, C.-T. Lin, and Z. Cao, "Deep neuro-cognitive co-evolution for fuzzy attribute reduction by quantum leaping PSO with nearest-neighbor memeplexes," *IEEE Trans. Cybern.*, vol. 49, no.7, pp. 2744-2757, Jul. 2019.

Weiping Ding (M'16-SM'19) received the Ph.D. degree in Computation Application, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2013. From 2014 to 2015, He is a Postdoctoral Researcher at the Brain Research Center, National Chiao Tung University (NCTU), Hsinchu, Taiwan. In 2016, He was a Visiting Scholar at National University of Singapore (NUS), Singapore. From 2017 to 2018, he was a Visiting Professor at University of Technology Sydney (UTS), Ultimo, NSW, Australia. He is a member of Senior IEEE and Senior CCF. His main research directions involve data mining, fuzzy systems, and big data analytics. He has published over 60 papers in flagship journals. Dr. Ding currently serves on the Editorial Advisory Board of *Knowledge-Based Systems*, and Editorial Board of *Information Fusion*, and *Applied Soft Computing*. He serves/served as an Associate Editor of several prestigious journals, including *IEEE Transactions on Fuzzy Systems*, and *Information Sciences*.

Witold Pedrycz (F'98) is Professor and Canada Research Chair (CRC) in Computational Intelligence in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. In 2009 Dr. Pedrycz was elected a foreign member of the Polish Academy of Sciences. In 2012 he was elected a Fellow of the Royal Society of Canada. He is a recipient of the IEEE Canada Computer Engineering Medal, and a Fuzzy Pioneer Award from the IEEE Computational Intelligence Society. His main research directions involve computational intelligence, fuzzy modelling, granular computing, and knowledge discovery. He has published numerous papers in these areas; the current h-index is 107 (Google Scholar) and 82 on the list *top-h scientists for computer science and electronics*. He is an Editor-in-Chief of *Information Sciences*, and Editor-in-Chief of *WIREs Data Mining and Knowledge Discovery* (Wiley).

Chin-Teng Lin (S'88-M'91-SM'99-F'05) received the B.S. degree from National Chiao-Tung University (NCTU), Taiwan in 1986, and the Master and Ph.D. degree in electrical engineering from Purdue University, USA in 1989 and 1992, respectively. He is currently the Distinguished Professor of Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. Dr. Lin was elevated to be an IEEE Fellow in 2005, and was elevated International Fuzzy Systems Association (IFSA) Fellow in 2012. Dr. Lin received the IEEE Fuzzy Systems Pioneer Awards in 2017. He served as the Editor-in-chief of IEEE Transactions on Fuzzy Systems from 2011 to 2016. He served as the Deputy Editor-in-Chief of IEEE Transactions on Circuits and Systems-II in 2006-2008. He has published more than 300 journal papers in the areas of neural networks, fuzzy systems, and brain computer interface, including over 120 IEEE journal papers.